YOUNG-GEUN KIM

Department of Biostatistics, Columbia University

Younggeun.Kim@nyspi.columbia.edu ⋄ Google Scholar

BIOGRAPHICAL INFORMATION

Name: Young-geun Kim
Date of Birth: September 10, 1991
Citizenship: Republic of Korea

JOB EXPERIENCE

Adjunct Associate Research Scientist   *Jul. 2021 - Present*
Department of Biostatistics, Columbia University
Research Scientist II   *Jul. 2021 - Present*
Research Foundation for Mental Hygiene, New York State Psychiatric Institute

Postdoctoral Researcher   *Mar. 2021 - Jun. 2021*
Department of Statistics, Seoul National University

EDUCATION

Seoul National University   *Mar. 2015 - Feb. 2021*
Ph.D. in Statistics   Graduated with the Best Dissertation Award
Advisor: Myunghee Cho Paik, Ph.D.
Dissertation: Statistical distance of conditional distributions and its applications

Seoul National University   *Mar. 2010 - Feb. 2015*
Triple Major   Graduated with Honors (Cum Laude)
B.S. in Industrial Engineering
B.S. in Statistics
B.S. in Mathematical Sciences

RESEARCH INTERESTS

My research interests include theoretical properties of statistical distances and their application to deep learning. Recently, I developed an interpretable deep representation learning algorithm for Adolescent Brain Cognitive Development [URL] study dataset, the largest single-cohort prospective longitudinal study of neurodevelopment and children's mental health in the United States.

Supervised learning and semi-supervised learning

- Classification and segmentation with imbalanced data
- Robust inference through kernel smoothing.

Unsupervised learning

- Anomaly detection
- Wasserstein generative models for incomplete and sequential data

Deep learning-based applications

- Interpretable representation learning for resting-state fMRI

- Classification and segmentation of intracranial hemorrhage in brain CT scans
- Network security and visual surveillance
- Video interpolation and extrapolation

HONORS AND AWARDS

Best Dissertation Award   *Feb. 2021*
College of Natural Sciences, Seoul National University

Seoul National University Innovation Program Scholarship   *Mar. 2017 - Feb. 2018*
Seoul National University
∗ Awarded to the Ph.D. student with the highest GPA in the department.

Student Paper Competition 1st Prize   *June 2017*
Korean Statistical Society

Brain Korea 21 Plus Scholarship   *Mar. 2016 - Feb. 2017, Mar. 2020 - Feb. 2021*
National Research Foundation of Korea

Merit-based Scholarship   *Mar. 2015 - Feb. 2016*
Seoul National University

National Scholarship for Science and Engineering   *Mar. 2010 - Feb. 2014*
Korea Student Aid Foundation

RESEARCH EXPERIENCE

I participated the following researches as a **research scientist**.

A data science framework for empirically evaluating and deriving   *Jul. 2021 - Present*
reproducible and transferrable RDoC constructs in youth (R01)
Funded by National Institutes of Health, U.S. Department of Health & Human Services

Computational approaches for validating dimensional constructs   *Jul. 2021 - Present*
of relevance to psychopathology (R01 clinical trial optional)
Funded by National Institutes of Health, U.S. Department of Health & Human Services

Deep learning with incomplete and sequential data: Application to *Mar. 2020 - Jun. 2021*
biomedical data
Funded by National Research Foundation of Korea

Development of low-yield trackers via causal inference   *May 2019 - Nov. 2019*
Funded by SK Telecom

Statistical approaches to deep learning: New methods for   *Mar. 2017 - Feb. 2020*
convolutional neural networks in application to medical
imaging data
Funded by National Research Foundation of Korea

Deep Learning for the CT based Acute Cerebral Infarction   *July 2016 - May 2019*
Classification and Lesion Segmentation
Collaborated with Seoul National University Bundang Hospital
Funded by National Research Foundation of Korea

New Robust Methods for Missing or Censored Covariates   *Mar. 2016 - Nov. 2016*
Funded by National Research Foundation of Korea

PEER-REVIEWED PUBLICATIONS

Journal

- **Kim, Y.-G.**, Lee, K., and Paik, M.C (2022). Conditional Wasserstein generator. *IEEE Transactions on Pattern Analysis and Machine Intelligence (Preprints)*. [Paper] [GitHub]
  - IF: 24.314 (**Top 2, upper <1%** on EE); h-index: 377 (**Top 1, upper <1%** on AI)

- **Kim, Y.-G.**, Kwon, Y., and Paik, M.C. (2019). Valid oversampling schemes to handle imbalance. *Pattern Recognition Letters*, 125 (1): 661-667. [Paper] [GitHub]
  - IF: 4.757 (**Top 53, upper 37%** on AI); h-index: 163 (**Top 12, upper 5%** on AI)

Conference

- Kim, M., **Kim, Y.-G.**, Kim, D., Kim, Y., and Paik, M.C. Kernel-convoluted deep neural networks with data augmentation. *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2021)*. [Paper] [GitHub]
  - IS: 32.10 (**Top 6, upper 3%** on ML and AI)

- **Kim, Y.-G.**, Kwon, Y., Chang, H., and Paik, M.C. (2020). Lipschitz continuous autoencoders in application to anomaly detection. *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics (AISTATS 2020)*. [Paper] [GitHub]
  - IS: 10.10 (**Top 18, upper 6%** on ML and AI)

WORKING PUBLICATIONS

Kim, Y.-G., Liu, Y., and Wei, X. Covariate-informed Representation Learning to Prevent Posterior Collapse of iVAE. [Paper]

Kim, Y.-G., Lee, K., Choi, Y., Won, J.-H., and Paik, M.C. Wasserstein geodesic generator for conditional distributions (under preparation).

Ravid, O.*, **Kim, Y.-G.***, Zhang, X., Kim, Y., Neria, Y., Wall, M., Lee, S., He, X., and Zhu, X. A visualization tool for variational autoencoder (under preparation).

Kim, Y.-G., Chang, W., Jung, S., and Paik, M.C. Few-shot Wasserstein translator (under preparation).

Kim, B.J., Choi, Y., **Kim, Y.-G.**, Paik, M.C., and Won, J.-H. Automated classification and segmentation of intracranial hemorrhage in brain CT scans with convolutional neural networks (under preparation).

∗: shared first author

PATENTS

Paik, M.C., **Kim, Y.-G.**, and Chang, H., Learning method and learning device for high-dimension unsupervised anomaly detection using kernalized Wasserstein autoencoder to lessen too many computations of Christophel function, and testing method and testing device using the same (KR102202842B1). [Patent]

Paik, M.C., **Kim, Y.-G.**, and Lee, K., Method and apparatus for conditional data generation using conditional Wasserstein generator (Submitted to Republic of Korea patent).

EVENT CHAIRING

Eastern North American Region (ENAR) 2023 Spring meeting (scheduled)
- Chair of the session "Advanced Methods for Analyzing Large-Scale Neuroimaging Data from Nationwide Consortiums for Mental Health Research" [Session Information]

International Conference on Machine Learning 2022
- Chair of the session "Theory" [Session Information]

INVITED PRESENTATIONS

International

- Kim, M., **Kim, Y.-G.**, Kim, D., Kim, Y., and Paik, M.C. (2021). Kernel-convoluted deep neural networks with data augmentation. *The 35th AAAI Conference on Artificial Intelligence (AAAI-21), Virtual conference due to COVID-19.*

- **Kim, Y.-G.**, Kwon, Y., Chang, H., and Paik, M.C. (2020). Lipschitz continuous autoencoders in application to anomaly detection. *The 23rd International Conference on Artificial Intelligence and Statistics (AISTATS 2020), Virtual conference due to COVID-19.*

- Kim, M., **Kim, Y.-G.**, Kim, D., Kim, Y., and Paik, M.C. (2020). Kernel-convoluted deep neural networks with data augmentation. *The 4th International Conference on Econometrics and Statistics (EcoSta 2020), Virtual conference due to COVID-19.*

- **Kim, Y.-G.**, Kwon, Y., Chang, H., and Paik, M.C. (2019). Lipschitz continuous autoencoders in application to anomaly detection. *IMS-China International Conference on Statistics and Probability, Dalian, China.*

- **Kim, Y.-G.**, Kwon, Y., and Paik, M.C. (2017). Handling imbalance in deep convolutional neural network: Application to medical imaging. *Neural Information Processing Systems 2017 (NIPS 2017) Workshop on Medical Imaging meets NIPS, Long Beach, CA, USA.*†

- Paik, M.C., Kwon, Y., and **Kim, Y.-G.**. (2017). Statistical approach to deep convolutional neural networks for medical imaging. *Data Science & Computational Precision Health 2017 (DahShu 2017), San Francisco, CA, USA.*

United States

- **Kim, Y.-G.**, Liu, Y, and Wei, X (2022). Covariate-informed Representation Learning with Identifiable Variational Autoencoders. *The 9th Annual Thomas R Ten Have Symposium on Statistics in Mental Health, NY.*†

Republic of Korea

- **Kim, Y.-G.**, Lee, K., and Paik, M.C. (2022). Conditional Wasserstein generator. *Spring Korea Statistical Conference 2022, Seoul.*

- **Kim, Y.-G.**, Chang, H., and Paik, M.C. (2018). Unsupervised anomaly detection using inverse Christoffel function via kernelized Wasserstein autoencoders. *Fall Korea Statistical Conference 2018, Seoul.*

- **Kim, Y.-G.**, Kwon, Y., and Paik, M.C. (2017). How to handle unbalanced dataset in medicine. *Deep learning educational workshop from basics to advances in medicine, Asan medical center, Songpa.*

† indicates a poster presentation.

CONTRIBUTED PRESENTATIONS

Kim, Y.-G., Kwon, Y., and Paik, M.C. (2017). Handling imbalance in medical imaging data using convolutional neural network. *Spring Korea Statistical Conference 2017, Seoul, Republic of Korea.*

OTHER ACADEMIC ACTIVITIES

Reviewer

- Journal

  - Expert Systems with Applications
    - IF: 8.665 (**Top 23, upper 8%** on EE); h-index: 225 (**Top 2, upper <2%** on AI)

  - Pattern Recognition Letters
    - IF: 4.757 (**Top 53, upper 37%** on AI); h-index: 163 (**Top 12, upper 5%** on AI)

  - International Journal of Computer Assisted Radiology and Surgery
    - IF: 3.421 (**Top 61, upper 29%** on Surgery); h-index: 53 (**Top 24, upper 23%** on Health Informatics)

- Conference

  - International Conference on Machine Learning (2022)
    - Selected as one of **Top 10%** of Reviewers [Outstanding Reviewer List]
    - IS: 32.40 (**Top 5, upper 2%** on ML and AI)

  - International Conference on Artificial Intelligence and Statistics (2022 and 2023)
    - IS: 10.10 (**Top 18, upper 6%** on ML and AI)

TEACHING EXPERIENCE

Guest Lecturer

- Deep Learning: A Statistical Perspective (Fall 2021)
  - Graduate-level course on deep learning.
  - Gave the lecture "Conditional Image Synthesis and Its Applications" in English.

Student Lecturer

- Deep Learning: A Statistical Perspective (Spring 2018, Fall 2018, Fall 2019, Fall 2020)
  - Graduate-level course on deep learning.
  - Gave lectures about deep learning programming languages and deep learning-based object detection algorithms in English.

- Seminar in Recent Developement of Applied Statistics (Fall 2017)
  - Graduate-level course on missing data analysis.
  - Gave a lecture about the application of expectation-maximization algorithm in incomplete data in English.

- Statistics Lab. (Fall 2015)
  - Freshman course to introduce R programming.
  - Gave whole lectures.

Teaching Assistant

I held office hours and graded homeworks and exams for the following courses.

- Mathematical Statistics 1 (Spring 2016, Summer 2016, Spring 2017, Summer 2017)
  - Major core course to focus on conditional probability, stochastic independence, and the distributions of random variables.

- Mathematical Statistics 2 (Fall 2016, Winter 2016, Fall 2017)
  - Major core course to provide a deeper understanding of limit distributions, statistical estimation, and statistical inferences.

- Statistics (Spring 2015, Spring 2020)
  - Freshman course to introduce Statistics.

TECHNICAL STRENGTHS

Programming Languages	Python, R, Matlab
Deep Learning	Tensorflow, Pytorch, Keras

LANGUAGE PROFICIENCY

Korean (Native), English (Fluent)